Journal of Harbin Engineering University Vol 44 No. 9
ISSN: 1006-7043 September 2023

A Deep Learning Fusion Approach for Mask Detection: CNN and VGG16 Integration

Heta S. Desai[1], Atul M. Gonsai[2]
SASCMA BCA College, Veer Narmad South Gujarat, University, Surat, 395017, Gujarat, India[1]
Department of Computer Science, Saurashtra University, Rajkot, 360005, Gujarat, India[2]

Abstract
Since the COVID-19 epidemic began, wearing a face mask has become a crucial precaution to stop the virus from spreading. In this research, we present a mask detection system that combines the VGG16 model with a Convolutional Neural Network (CNN) architecture. To do this, we construct a comprehensive dataset image of people with and without masks set against diverse backgrounds. The training, validation, and testing sets are then created from the dataset. The pre-trained VGG16 model is utilized as a feature extractor to pull out distinctive qualities from the input images. The outcomes show how well the fusion of CNN with VGG16 model can distinguish between masked and unmasked people even in difficult situations involving occlusions and a wide range of backdrops. We demonstrate the proposed method's better accuracy and computational effectiveness by comparing it to state-of-the-art mask detection approaches. The system is a trustworthy tool for mask detection in situations in the real-world including airports, hospitals, and public areas since it achieves an overall accuracy of over 99.47% of training and 98.13% of validation respectively.
Keywords-Mask Detection, CNN, VGG16, Occlusion

1. Introduction

Since the COVID-19 epidemic, masked-based face detection has become more and more necessary to verify compliance with mask regulations and preserve public health [1]. There are numerous parameters that can affect the outcome of any algorithm when a face is detected from a surveillance camera, including face angle, facial expression variation, illumination, camera angle, occlusion, ageing, and many more. It may be difficult to identify people wearing masks using conventional facial detection techniques, making it difficult to monitor and enforce mask-wearing regulations [2]. The choice of algorithm will depend on the particular needs of the application, such as speed, accuracy, and the availability of training data. However, there are many existing algorithms Viola Jones, CNN, HOG, SIFT [3] works on mask detection with specifically designed environments. But there are rare algorithms which solve the problem of face mask detection in natural environments [4].

Mask-based face detection is a method for spotting people wearing masks in public spaces or other areas where it's appropriate to cover a person's face. Mask wearing has become widespread since the COVID-19 outbreak, making it difficult to detect people using conventional face detection techniques. In order to identify people based on facial features that are still discernible when wearing a mask, such as the eyes and forehead, mask-based face detection methods use machine learning and computer vision techniques. Mask-based face detection can assist increase public safety and strengthen security measures in a variety of applications, including security, transportation, and medical care, by making accurate face detection possible in circumstances when masks are necessary.

The development of a reliable mask detection system using fusion of CNN and VGG16 architecture, the creation of a sizable dataset, and the performance testing of the suggested model are all contributions of this study. The results show how deep learning techniques may be used to solve problems in public health and make it easier to put preventive measures in place during pandemics.

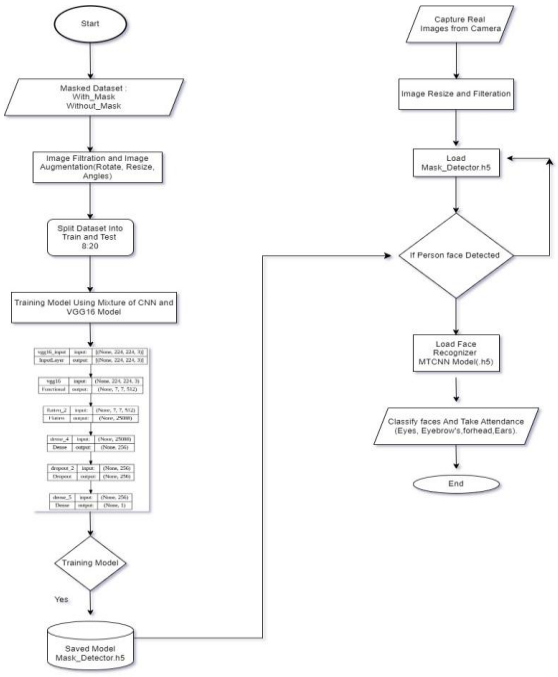

Figure 1 System Flow Chart

The whole mask-based face recognition process using surveillance cameras is depicted in the above figure. Face detection, Mask detection, and Face recognition system make up this mask-based face recognition system. Videos are initially captured using surveillance cameras, after which frames are retrieved, and a face identification algorithm is subsequently applied to those frames. The Haar Cascade technique is what we are employing for the face detection process. After utilising the Haar Cascade to identify the face, the Mask detection technique was used. We developed a model which is fusion of Convolutional Neural Network and VGG16 model for the aim of mask detection to determine whether or not the person is wearing a face mask. Face recognition is the last step after mask detection.

2. Related Work

A novel approach, light-CNN, based on the VGG16 model was proposed by Anugrah Perdana and Adhi Prahara [5], but it had a small dataset. Comparing the suggested approach to the conventional CNN model, the number of stages has been decreased. Researchers have completely deleted 256 and 512 layers as well as one layer from 64 filters in order to create a lightweight and compact-CNN based on VGG16. The dimensions of the fully interconnected layers have also changed. Also, they contrasted the typical VGG16 model with the suggested model and achieved 94.4% accuracy for a dataset with few labels.

Using three phases, Bendjillali Ridha Ilyas [6] introduced a novel face recognition method. For face detection, they employed the Viola Jones method for detection purpose, for facial image enhancement they used the Adaptive Histogram Equalization algorithm, and for classification they used the VGG16 and ResNet50 Neural Network. The proposed approach was tested using the CMU PIE dataset and the expanded Yale B dataset. Researchers compared the suggested model on both datasets and found that VGG16 and ResNet50 had accuracy of 96.12% and 97.23% on the Yale B dataset and 96.55% and 98.28% on the CMU PIE Dataset, respectively.

[3] Htet Aung the VGG16 Pre-trained CNN model and the YOLO algorithm were used to improve the face identification system. For the face detection system, the 5000-image FDDB standard picture dataset was employed. This dataset contains photographs with diverse skin tones, expressions, and poses. They compared the YOLO algorithm with two distinct models, including the YOLO+ Alex Network model and the YOLOv2+ Google Network Model, along with the proposed YOLOv2+VGG16 model, which was 93% accurate.

A new technique that uses a CNN and VGG16-based deep learning algorithm was proposed by Alok Negi and Krishan Kumar [1]. This strategy aids in locating the individual in a crowd who is not covering their face with face masj. Five levels of maxpooling and two layers of flattening were employed in the suggested model. Batch size 32 and the Adam Optimizer have been utilized. The SMFD dataset has been used for validation, training, and testing. In terms of training, validation, and testing accuracy, the CNN technique achieves 96.35%, 96.35%, and 97.42%, while the VGG!6 model achieves 99.47%, 98.59%, and 98.97%, respectively.

A system developed by Chamandeep Vimal and Neeraj Shrivastava [7] allows to obtain greater accuracy when faces are covered and hidden by objects, which is not possible with typical face detection systems. CNN-based VGG16 has been utilized with a few shallow layers that can function in a real-world setting. The classification of photos

as masked or unmasked has been done using the pre-trained VGG16 model. This model's suggested detection accuracy is 93%.

Adhitya Velip and Amita Dessai have created a novel system [2] to determine whether or not someone is wearing a face mask. VGG16, MobileNetV2, and Dense121, a CNN-based model, have been used to measure a system's performance. For feature extraction, the CNN-based models VGG16, MobileNetV2, and DenseNet121 were applied. Set the input image size to 224 X 224 X 3 and 15 epochs for testing purposes. The training accuracy for VGG16, MobileNetv2, and DenseNet121, respectively, was determined to be 93.1%, 99.88%, and 99.37%.

An autonomous method for face identification and recognition utilizing Eigenfaces was proposed by Susetyo Bagas Bhaskoro, Siti Aminah, and Khoutal Taqi [8]. In order to verify the accuracy result in terms of varying light levels and object distance, VGG16 based on CNN architecture has been applied. The high error rate in our proposed system is caused by things that are covered by face masks or other concealing devices. Each of the dataset's ten classes has 40 training records and 10 validation records, respectively. System achieves training and validation accuracy of 85.72% and 96.25%, respectively.

The researchers' primary goal is to submit a work to create face detection in illumination conditions, according to Ivana Lucia Kharisma and Rahmadya Trias Handayanto [9]. In order to implement this work, CNN and pre-trained models like VGG16 and VGG19 were used. This hypothetical model determines whether a face mask is worn or not. 800 total photos were divided into mask and no mask categories in the dataset. The training and validation accuracy for CNN is 97.79% and 97.22%, VGG16 is 99.87% and 99.00%, and VGG19 is 100% and 99.00%. Further extensions of this work will include third-class inappropriate facemasks, which can be used to determine if a facemask is worn above or below the chin.

The study on various algorithms, including YOLO, SSD, CNN, VGG16, MobileNet, MTCNN, ResNet, RCNN, and Viola Jones, was presented in this research work by RakhsithL.A, AnushaK.S, Karthik B.E, ArunNithish D, and KishoreKumarV [10]. Of all of these methods, the YOLO algorithm was shown to have the highest accuracy and least amount of loss. Viola Jones has achieved 90% accuracy but it works well if all of the eyes are visible. VGG16 has achieved 92.13% accuracy but with minimal dataset. Tensorflow with CNN has reached 98.9% validation accuracy with a little dataset, whereas MobileNetv2 has obtained 97.11% accuracy with little data.

3.Methodology

3.1Data Collection

There are a total almost 2000 photos of human faces with and without masks in the training, validation and testing archive. In this research, our goal is to develop a reliable image classification model with constraints that only include a small number of training examples of both mask- and non-masked face images. To impose a constraint on the input photographs, we must reduce the number of images. We selected 80% of the photos for training and 20% for Testing from the dataset.

Figure 2 Sample of dataset images

3.2 Selected Model

This proposed algorithm combines the VGG16 Model which serves as feature-extractor, with a custom-built CNN classifier for mask detection. The VGG16 model is a pre-trained model, which implies that before being made available for usage, it underwent training on a sizable dataset of images. To minimize the discrepancy between the projected output and the actual output, the weight of neural networks is iteratively rearranged while the model is being trained with a group of images.

For image identification tasks, this pre-trained VGG16 model can be used as a starting point. Then Its performance can be enhanced using a particular dataset. Aside from object detection, segmentation, and localization, the VGG16 model's pre-trained weights can be used to extract

features from photos for use as input in other machine learning models.

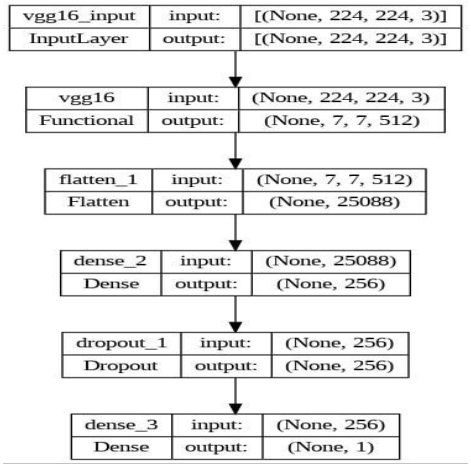

Figure 3 Proposed Model Architecture

The proposed model uses pre-trained weights from the dataset and the convolutional neural network architecture VGG16. The top layers of the VGG16 model are not included since the include_top option has been set to FALSE. Additionally, input_shape may be set to input_shape to set the input shape. The VGG16 model's layers are frozen by setting ea ch layer's trainable attribute to False using a for lo op in

order to avoid weight updates during training. The Keras Sequential API is then used to construct the model. By using the add function on the model, the VGG16 model has been added as the top layer. The output of the convolutional layers is then transformed into a 1-dimensional feature vector by adding a Flatten layer. Then, to lessen overfitting, a Dropout layer with a rate of 0.5 is inserted after a Dense layer with 256 hidden nodes and the ReLU activation function. For binary classification, a Dense layer with a single output node and the sigmoid activation function is added. A multi-layer neural network for binary classification using pre-trained weights and frozen convolutional layers from the VGG16 model makes up the resultant architecture.

The pre-trained VGG16 model is used as a feature extractor in the VGG16+CNN architecture, and it is combined with a unique CNN classifier for mask detection. By fine-tuning the model on a mask detection dataset, it learns to distinguish between faces with and without masks. This combined architecture offers an effective solution for automated mask detection, contributing to various applications related to public health and safety.

4.Result Analysis

As stated previously, we'll begin by creating a basic convolutional neural network from scratch, train it using a training image dataset, and then assess the outcomes. Later, to increase the accuracy, we'll employ an image augmentation technique. Finally, we will extract features and categorize photos using the pre-trained model VGG16, which was previously trained on a customized dataset with a wide range of categories.

The CNN model is used to process the input images. Convolutional neural networks and the pre-trained VGG16 model were used to train the dataset. There are two sections to the dataset. 20% of the images are utilized for testing and 80% are used for training. Results of the training process are presented in the form of validation, loss accuracy, and loss for every epoch. Figures 5 and 6 respectively represent the accuracy and loss plots for the model, namely the fusion of CNN and VGG16 model.

Table 1 describes training and validation accuracy and loss for every epoch. Proposed model accuracy for training and validation has been reached at 99.31% and 97.66%, respectively.

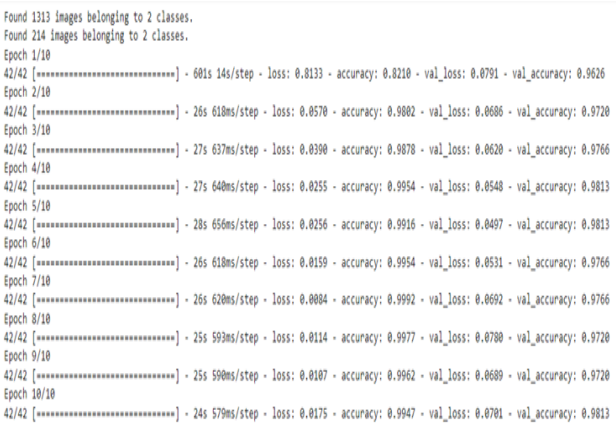

Figure 4 Training Process for 10 epochs

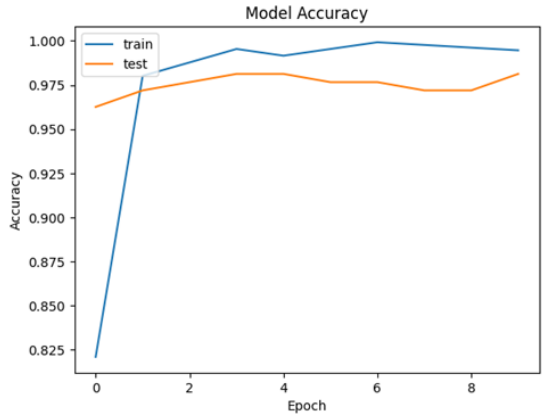

Figure 5 Training and validation Accuracy

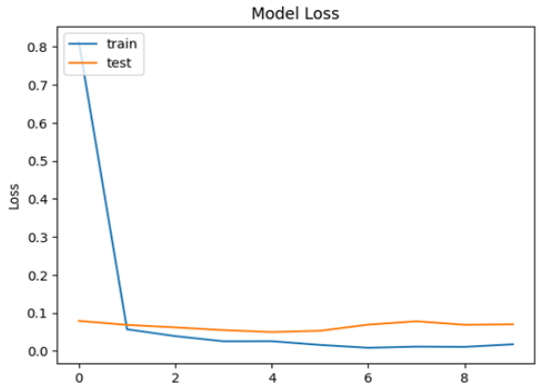

Figure 6 Training and validation Loss

Table 1 Training Process for 10 epochs

S. No	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.8133	0.8210	0.0791	0.9626
2	0.0570	0.9802	0.0686	0.9720
3	0.0390	0.9878	0.0620	0.9766
4	0.0255	0.9954	0.0548	0.9813
5	0.0256	0.9916	0.0497	0.9813
6	0.0159	0.9954	0.0531	0.9766
7	0.0084	0.9992	0.0692	0.9766
8	0.0114	0.9977	0.0780	0.9720
9	0.0107	0.9962	0.0689	0.9720
10	0.0175	0.9947	0.0701	0.9813

When compared to other algorithms, the proposed algorithm has the closest 99% training accuracy and 97% validation accuracy. Comparison between the proposed and method is shown in Table 2.

Table 2 Result Analysis with another existing algorithm.

S. No	Algorithm	Result	Future Scope
1	SSD- mask algorithm [11]	90.2% Precision and 86.5% recall.	High Detection accuracy can be achieved.
2	VGG16 architecture based on CNN [7]	95% training and 93% validation accuracy on small dataset.	Will work on large dataset and work with better camera with deem light.
3	MobileNetV2 [12]	99.72% and 99.82% of training and validation accuracy respectively.	Improve to detect humans without mask automatically.
4	CNN and VGG16 model on SMFD dataset [1]	CNN archives 96.35% and 97.42% training and validation accuracy. VGG16 achieves 99.47% and 98.59% of training and validation accuracy.	Will use video surveillances for capturing live feed.
5	YOLOV5 [13]	Maximum mAP of model with 88.1% of detection accuracy.	Will increase accuracy of detection by applying other deep learning technique.
6	InceptionV3 + LR [14]	Obtain 96% of detection accuracy.	Will use high resolution camera and extend to recognize the face.
7	CNN +	97% of	Can be used

	improved AlexNet + improve VGG16 [15]	average detection accuracy.	by companies and organization.
8	InceptionV3 [16]	96.77% of training and 89.29% of Validation Accuracy and 0.0997 0.4906 training and validation loss.	It takes longer amount of time.
9	Improvised VGG16 model. [17]	Applied on CK+ and JAFFE dataset and achieves 94.8% and 93.7% accuracy respectively.	Will freeze the intermediate layers of VGG16 model on same dataset.
10	Yolo Nano based face mask detection [18].	Yolo Nano Object Detection model. L2 function is used for mask detection.	Will work on expression recognition for face mask detection.
11	Proposed model:	99.47% of training and 98.13% of validation accuracy.	Lower Resolution Camera can decrease the accuracy.

Conclusion

In conclusion, the utilization of a fusion of CNN with VGG16 model for mask detection has proven to be highly effective. The combination of deep learning techniques and the VGG16 architecture has enabled accurate identification of masked and unmasked faces. This study aids in the creation of real-time apps for mask identification in various settings. The results demonstrate the model's robustness and potential for mitigating the spread of contagious diseases. The model achieves 99.47% and 98.13% of training and validation accuracy respectively. Further advancements in this area will continue to enhance public health and safety measures. Future work on another dataset will be done to test accuracy, and this portion can be integrated into the facial recognition phase.

*Author for correspondence

References

[1] Negi A, Kumar K, Chauhan P, Rajput RS. Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic. In2021 international conference on computing, communication, and intelligent systems (ICCCIS) 2021 Feb 19 (pp. 595-600). IEEE.

[2] Velip A, Dessai A. Face Mask Detection Using Machine Learning Techniques. In2022 2nd Asian Conference on Innovation in Technology (ASIANCON) 2022 Aug 26 (pp. 1-5). IEEE.

[3] Aung H, Bobkov AV, Tun NL. Face detection in real time live video using yolo algorithm based on Vgg16 convolutional neural network. In2021 International conference on industrial engineering, applications and manufacturing (ICIEAM) 2021 May 17 (pp. 697-702). IEEE

[4] Cheng X, Wang J, Goh S. Enhancing YOLOv3-tiny for Mask Detection in Natural Scenes. In2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC) 2022 Dec 2 (pp. 1073-1077). IEEE.

[5] Perdana AB, Prahara A. Face recognition using light-convolutional neural networks based on modified Vgg16 model. In2019 International Conference of Computer Science and Information Technology (ICoSNIKOM) 2019 Nov 28 (pp. 1-4). IEEE.

[6] Ilyas BR, Mohammed B, Khaled M, Miloud K. Enhanced face recognition system based on deep CNN. In2019 6th International Conference on Image and Signal Processing and their Applications (ISPA) 2019 Nov 24 (pp. 1-6). IEEE.

[7] Vimal C, Shirivastava N. Face and Face-mask Detection System using VGG-16 Architecture based on Convolutional Neural Network. International Journal of Computer Applications.;975:8887.

[8] Bhaskoro SB, Aminah S, Taqi K. Attendance System on Moving Objects through Face Recognition using MTCNN and CNN. In2021 3rd International Symposium on Material and Electrical Engineering Conference (ISMEE) 2021 Nov 10 (pp. 184-189). IEEE.

[9] Kharisma IL, Handayanto RT, Dewi DA. Face Mask Detection In The Covid-19 Pandemic Era by Implementing Convolutional Neural Network and Pre-Trained CNN Models. In2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED) 2021 Aug 5 (pp. 1-6). IEEE.

[10] Rakhsith LA, Anusha KS, Karthik BE, Nithish DA, Kumar VK. A survey on object detection methods in deep learning. InProc. of 2021 Second Int. Conf. on Electronics and Sustainable Communication Systems (ICESC) 2021.

[11] Xu M, Wang H, Yang S, Li R. Mask wearing detection method based on SSD-Mask algorithm. In2020 International Conference on Computer Science and Management Technology (ICCSMT) 2020 Nov 20 (pp. 138-143). IEEE.

[12] Shamrat FJ, Chakraborty S, Billah MM, Al Jubair M, Islam MS, Ranjan R. Face Mask Detection using Convolutional Neural Network (CNN) to reduce the spread of COVID-19. In2021 5th international conference on trends in electronics and informatics (ICOEI) 2021 Jun 3 (pp. 1231-1237). IEEE.

[13] Yin J, Jin J. A Face Mask Detection Algorithm Based on YOLOv5. In2022 International Conference on Machine Learning, Control, and Robotics (MLCR) 2022 Oct 29 (pp. 55-61). IEEE.

[14] Reddy S, Goel S, Nijhawan R. Real-time face mask detection using machine learning/deep feature-based classifiers for face mask recognition. In2021 IEEE Bombay Section Signature Conference (IBSSC) 2021 Nov 18 (pp. 1-6). IEEE.

[15] Song Z, Nguyen K, Nguyen T, Cho C, Gao J. Camera-based security check for face mask detection using deep learning. In2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService) 2021 Aug 23 (pp. 96-106). IEEE.

[16] Poyekar B, Mote R, Shah J, Dholay S. Face Recognition Attendance System for Online Classes. In2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) 2022 Oct 3 (pp. 1-6). IEEE.

[17] Dubey AK, Jain V. Automatic facial recognition using VGG16 based transfer learning model. Journal of Information and Optimization Sciences. 2020 Oct 2;41(7):1589-96.

[18] Nasiri E, Milanova M, Nasiri A. Video Surveillance Framework Based on Real-Time Face Mask Detection and Recognition. In2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA) 2021 Aug 25 (pp. 1-7). IEEE.

Heta S. Desai is a Ph.D Scholar at department of Computer Science, Saurashtra University, Raajkot, Gujarat. She has qualified Gujarat State Eligibility Test (GSET) in computer science in 2018. She has received a Bachelor's degree in computer science from Veer Narmad South Gujarat University, Surat,Gujarat and Masters in Computer Science from Gujarat Technological Univerisity, Gujarat. She has more than 9 years of Teaching Experienece and Currently working as an assistant professor in SASCMA BCA college,Suart, Gujarat.

Email: prof.hetadesai@gmail.com

Dr. Atul M. Gonsai is a professor at computer science department, Saurashtra University, Rajkot, Gujarat. He has received many awards throughout his acedemic journey. He has published many patents as well as numerous research paper also. His intrested area of research is Computer Networking, Wireless Network, Network Security, Image feature extraction and classification. He is NAAC/AAA team member, ICT Resolution Committee member, UGC expert Committee member, member of CMI, IACSIT, UACEE, Fellow member for IETE, ACM, Life member for ISTE, CSI, ISCA also. He is BoS member of many recognized University and member of Advisory Body also. He is Coorinator and Nodal Officer of saurashtra University.

Email: amg@sauuni.ac.in